

A new stochastic process to model Heart Rate series during exhaustive run and an estimator of its fractality parameter

Jean-Marc Bardet

Université Paris 1, SAMOS-MATISSE-CES, 90 rue de Tolbiac, 75013 Paris, France.

Véronique Billat

Université d'Evry, LEPHE, E.A. 3872 Genopole, Boulevard F. Mitterrand, Evry Cedex, France.

Imen Kammoun

Université Paris 1, SAMOS-MATISSE-CES, 90 rue de Tolbiac, 75013 Paris, France.

Summary. In order to interpret and explain the physiological signal behaviors, it can be interesting to find some constants among the fluctuations of these data during all the effort or during different stages of the race (which can be detected using a change points detection method). Several recent papers have proposed the long-range dependence (Hurst) parameter as such a constant. However, their results induce two main problems. Firstly, DFA method is usually applied for estimating this parameter. Clearly, such a method does not provide the most efficient estimator and moreover it is not at all robust even in the case of smooth trends. Secondly, this method often gives estimated Hurst parameters larger than 1, which is the larger possible value for long memory stationary processes. In this article we propose solutions for both these problems and we define a new model allowing such estimated parameters. On the one hand, a wavelet-base estimator is applied to data. Such an estimator provides optimal convergence rates in a semiparametric context and can be used for smoothly trended processes. On the other hand, a new semiparametric model so-called locally fractional Gaussian noise is introduced and is characterized by a so-called parameter which can be larger than 1. Such semiparametric process is tested to be relevant for modeling HR data in the three characteristic phases of the race. It also shows an evolution of the local fractality parameter during the race confirming the results obtained by Peng *et al.* (1995) in their study regarding Hurst parameter of HR time series during the exercise for healthy adults (where the estimated parameter is close to that observed in the race beginning) and heart failure adults (where the estimated parameter is close to that observed in the end of race). So, this evolution, which can not be observed with DFA method, may be associated with fatigue appearing during the last phase of the marathon.

Keywords: Wavelet analysis; Detrended fluctuation analysis; Fractional Gaussian noise; Self-similarity; Hurst parameter; Long-range dependence processes; Heart rate time series

1. Introduction

The content of this article was motivated by a general study of physiological signals of runners recorded during endurance races as marathons. More precisely, after different signal procedures for "cleaning" data, one considers the time series resulting of the evolution of

E-mail: bardet@univ-paris1.fr

heart rate (HR) data during the race. The following figure provides several examples of such data (recorded during Marathon of Paris). For each runner, the periods (in ms) between the successive pulsations (see Fig. 1) are recorded. The HR signal in number of beats per minute (bpm) is then deduced (the HR average for the whole sample is of 162 bpm).

This paper, focuses on the modeling and the estimation of relevant parameters characterizing these instantaneous heart rate signals of athletes recorded during the marathon. We have chosen to focus in an exponent that can be called "Fractal", which indicates the local regularity of the path and the dependency between data. In certain stationary cases, this parameter is close to the Hurst parameter, defined for long range dependent (LRD) processes.

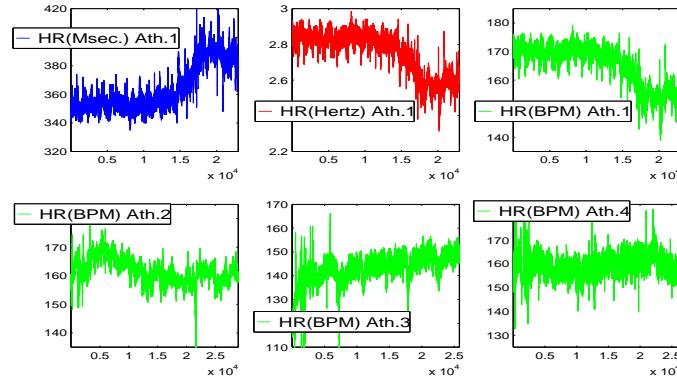


Fig. 1. Heart rate signals of Athlete 1 in ms, Hertz and BPM (up), of Athletes 2, 3 and 4 in BPM (down)

The LRD behavior is often seen on various data. This phenomenon was developed in many fields beginning with hydrology (Hurst, 1951), telecommunication, biomechanics and recently in economy and finance. A mathematical and signal processing methods have also noted the presence of LRD in time series describing the fluctuations over time of physiological signals (Goldberger (2001), Goldberger *et al.* (2002)). Indeed, numerous authors have studied heartbeat time series (see for instance Peng *et al.* (1993), (1995) or Absil *et al.* (1999)) and the model proposed to fit these data is a trended long memory process with an estimated Hurst parameter close to 1 (and sometimes more than 1). In this article, three improvements have been proposed to such a model:

- (a) Data are stepped in three different stages which are detected using a change points detection method (see for instance Lavielle (1999)) developed in Section 2. The main idea of the detection's method is to consider that the signal distribution depends on a vector of unknown characteristic parameters constituted by the mean and the variance. The different stages (beginning, middle and end of the race) and therefore the different vectors of parameters, which change at two unknown instants, are estimated. This first step is important since the model defined below fits well for each sub-series but not at all for the whole HR data.
- (b) The parameter H which is very interesting for interpreting and explaining the physiological signal behaviors, is estimating using two methods the DFA method and wavelet

analysis in Section 3.

The DFA which is a version, for time series with trend, of the aggregated variance method was studied in details in Bardet and Kammoun (2008). In particular, the asymptotic properties of the DFA function and the deduced estimator of H are studied in the case of fractional Gaussian noise and extended to a general class of stationary semiparametric long-range dependent processes with or without trend. We have shown that DFA method is not as efficient to estimate the Hurst parameter of stationary long memory processes, than other methods such as log-periodogram (see for instance, Moulines and Soulier (2003) or wavelet analysis (see Abry *et al.* (1998), or Moulines *et al.* (2007)), which provide the optimal convergence rate (in sense of the minimax criterium). Moreover, if the DFA method is not at all robust in the case of polynomial trends, this is not such a case of the wavelet analysis method. Finally, a goodness-of-fit test can be deduce from the wavelet analysis method. Despite the popularity of DFA method in numerous papers concerning such physiological signals (see for instance, Absil *et al.* (1999), Ivanov *et al.* (2001), Peng *et al.* (1993), (1995)), it is therefore clearly more interesting to use the wavelet based estimator in view of estimate a “fractal” exponent of HR data.

As a first model, the usual fractional Gaussian noise (FGN) is then proposed for modeling HR data. In such a context, the wavelet based estimator provides two results. Firstly, the estimated parameter often exceeds the value 1, which is the largest possible value for a FGN. Secondly, even for the 3 different stages of the race, the goodness-of-fit test is always rejected.

- (c) In Section 4, we propose to model HR data, during each stage, with a generalization of fractional Gaussian noise, called locally fractional Gaussian noise. Such stationary process is built from a parameter called local fractality parameter which is a kind of Hurst parameter in a restricted band frequency (that may take values in \mathbb{R} and not only in $(0, 1)$ as usual Hurst parameter). The estimation of local fractality parameter and also the construction of goodness-of-fit test can be made with wavelet analysis. We also show the relevance of model and an evolution of the parameter during the race, which confirms results obtained by other authors in their study regarding the distinguish of healthy from pathologic data (see Peng *et al.* (1995)).

2. Abrupt change detection

During effort, one or more phases can be observed in recorded HR series, which evolve and change differently from an athlete to another: the transition step, recorded between the race beginning and the stage of HR reached during the effort, the main stage during the exercise and an arrival phase until the race end. So, after cleaning the HR data (implying that only 9 athlete HR data are now considered) an automatic detection of changes is applied to HR time series cutting its in different race phases - beginning, middle and end.

The change point detection method used here is developed by Lavielle (see for instance Lavielle (1999)). The main idea is to consider that the signal distribution depends on a vector of unknown characteristic parameters in each stage. The different stages and therefore the different vectors of parameters, change at unknown instants. For instance and it will be our choice, changes in mean and variance can be detected. Applied to the data, the change point detection method along these two phenomena distinguishes beginning, middle

and end of race. So, it may be possible to envisage a piecewise stationarity *i.e.* that the process is almost stationary on fixed time intervals and it remains the modeling of this stationary component.

General principle of the method of change detection

Assume that a sample of a time series $(Y(i), i = 1, \dots, n)$ is observed. Assume also that it exists $\tau = (\tau_1, \tau_2, \dots, \tau_{K-1})$ with $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n = \tau_K$ and such that for each $j \in \{1, 2, \dots, K\}$, the distribution law of $Y(i)$ is depending on a parameter $\theta_j \in \Theta \subset \mathbb{R}^d$ (with $d \in \mathbb{N}$) for all $\tau_{j-1} < i \leq \tau_j$. Therefore, K is the number of segments to be deduced starting from the series and $\tau = (\tau_1, \tau_2, \dots, \tau_{K-1})$ is the ordered change instants. Now, define a contrast function

$$U_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})),$$

of $\theta \in \mathbb{R}^d$ applied on each vector $(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1}))$ for all $j \in \{0, 2, \dots, K-1\}$. A general example of such a contrast function is

$$U_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})) = -2 \log L_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})),$$

where L_θ is the likelihood. Then, for all $j \in \{0, 2, \dots, K-1\}$, define:

$$\widehat{\theta}_j = \operatorname{Argmin}_{\theta \in \Theta} U_\theta(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1})).$$

Now, set:

$$\widehat{G}(\tau_1, \dots, \tau_{K-1}) = \sum_{j=0}^{K-1} U_{\widehat{\theta}_j}(Y(\tau_j + 1), Y(\tau_j + 2), \dots, Y(\tau_{j+1}))$$

As a consequence, an estimator $(\widehat{\tau}_1, \dots, \widehat{\tau}_{K-1})$ can be defined as:

$$(\widehat{\tau}_1, \dots, \widehat{\tau}_{K-1}) = \operatorname{Argmin}_{0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < n} \widehat{G}(\tau_1, \dots, \tau_{K-1}). \quad (1)$$

The principle of such method of estimation is very general (it can be also devoted to estimate abrupt change in polynomial trends) and different asymptotic behavior of the estimator $(\widehat{\tau}_1, \dots, \widehat{\tau}_{K-1})$ can be deduced under general assumption on the time series Y (see for instance Bai Perron, 1998, Lavielle, 1999, or Moulines and Lavielle, 2000).

For HR data, it is obvious that the beginning and the end of the race implies respectively an increasing (respectively decreasing) of the mean of HR. However, for avoiding all confusion linked for instance to the stress of the runner or other harmful noises, it was chosen to detect a change in mean and variance.

Change detection in mean and variance

Therefore, for all $j \in \{0, 1, \dots, K-1\}$, consider the following general model:

$$Y(i) = \mu_j + \sigma_j \varepsilon_i \quad \text{for all } i \in \{\tau_j + 1, \dots, \tau_{j+1}\},$$

where $\theta_j = (m_j, \sigma_j) \in \mathbb{R} \times (0, \infty)$ and (ε_i) is a sequence of zero-mean random variables with unit variance.

In the case of changes in both mean and variance, and it is such a framework we consider for the heart rates series, a "natural" contrast function is defined by:

$$U_{\theta_j}(Y(\tau_j + 1), \dots, Y(\tau_{j+1})) = \sum_{\ell=\tau_j+1}^{\tau_{j+1}} \frac{(Y(\ell) - m_j)^2}{\sigma_j^2},$$

and therefore the well-known estimator of θ_j is:

$$\hat{\theta}_j = (\hat{m}_j, \hat{\sigma}_j) = \left(\frac{1}{\tau_{j+1} - \tau_j} \sum_{\ell=\tau_j+1}^{\tau_{j+1}} Y(\ell), \frac{1}{\tau_{j+1} - \tau_j} \sum_{\ell=\tau_j+1}^{\tau_{j+1}} (Y(\ell) - \hat{m}_j)^2 \right).$$

Now, the estimator $(\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$ can be deduced from (1). When the number of changes is unknown, the procedure is exactly the same except that the number of changes is estimated. Thus, a new contrast V is built by adding to the previous contrast U an increasing function depending on the change number K , *i.e.* more precisely,

$$\hat{V}(\tau_1, \dots, \tau_{K-1}, K) = \hat{G}(\tau_1, \dots, \tau_{K-1}) + \beta \times \text{pen}(K),$$

with $\beta > 0$. As a consequence, by minimizing V in $\tau_1, \dots, \tau_{K-1}, K$, an estimator \hat{K} is obtained which varies with the penalization parameter β .

For HR data, the choice of $\text{pen}(K)$ was K . Let $\hat{G}_K = \hat{G}(\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$, for $K = K_1, \dots, K_{MAX}$ we define

$$\beta_i = \frac{\hat{G}_{K_i} - \hat{G}_{K_{i+1}}}{K_{i+1} - K_i} \quad \text{and} \quad l_i = \beta_i - \beta_{i+1} \quad \text{with } i \geq 1.$$

Then the retained K is the greatest value of K_i such that $l_i \gg l_j$ for $j > i$.

Applied to the whole set HR data, the number of abrupt changes is estimated at 4 or 3. Three phases were selected to be studied, which are located in the beginning of the race, in the middle and in the end (see for example Fig. 2). However for certain recorded signals the first or the last phase can not be distinguished probably for measurements reasons.

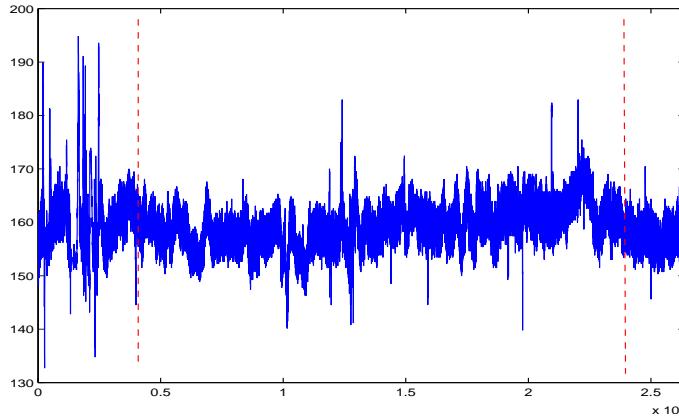


Fig. 2. The estimated configuration of changes in a HR time series of an athlete

In order to unveil if a change of behavior of HR series was happened during these three detected phases of the marathon, we propose a new model for these sub-series characterized by a parameter H . Several common estimators of this parameter, so-called scaling behavior exponents, consist in performing a linear regression fit of a scale-dependent quantity versus the scale in a logarithmic representation. This includes the Detrended Fluctuation Analysis (DFA) method Peng *et al.* (1994) and the wavelet analysis method Abry *et al.* (2003).

3. A fractional Gaussian noise for modeling HR series and the estimation of the Hurst parameter

In this section, a first model, the fractional Gaussian noise, is proposed for modeling HR data. After a statistic study, one chooses to estimate the Hurst parameter with a wavelet based estimator instead of the DFA method (which is commonly used in physiologic papers despite its weak performances). Moreover, a test built from the wavelet based method shows the badness-of-fit of this model to the data.

3.1. A first model: the fractional Gaussian noise

When we observe entire or partial (during the three phases) HR time series, we remark that it exhibits a certain persistence and the related correlations decays very slowly with time what characterizes trajectories of a long memory Gaussian noise. Also, the distribution of data recorded during the phases leads as to suspect a Gaussian behavior in these data. Of course this is only an assumption and we can check it with tests considered for long range dependent processes. But in our case we will try to test whether a Gaussian process could model these data. Moreover, the aggregated signals (see for example Fig. 3) present a certain regularity very close to that of fractional Brownian motion simulated trajectories with a parameter close to 1 (Fig. 5). So, fractional Gaussian noise could be an appropriated model to HR series.

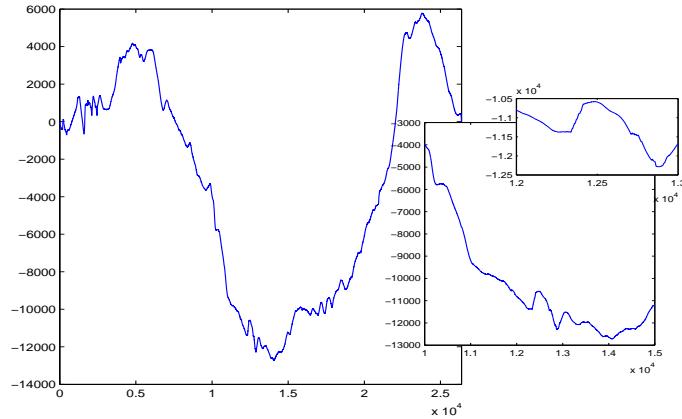


Fig. 3. The self-similarity of the aggregated HR signals (representation of the aggregated HR fluctuations at 3 different time resolutions)

The following Figure 4 presents a comparison between the graphs of HR data during a stage (detected previously) and a fractional Gaussian noise (FGN in the sequel) with parameter $H = 0.99$ (see the definition above). Before using statistical tools for testing the similarities of both these graphs, let us remind some elements concerning the FGN.

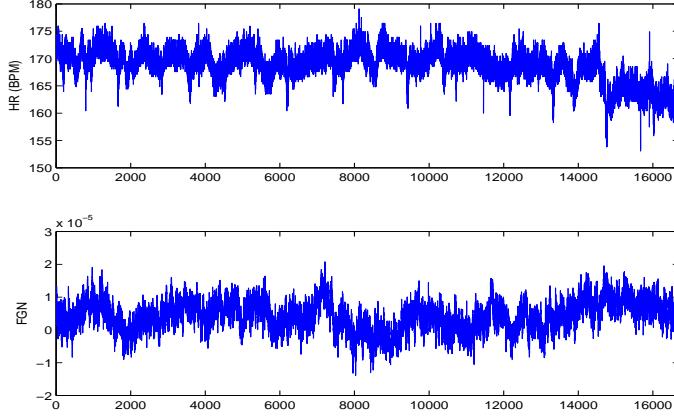


Fig. 4. Comparison of HR data in the middle of race (Ath4) and generated FGN($H=0.99$) trajectories

The FGN is one of the most famous example of stationary long range dependent (LRD in the sequel) process. The LRD phenomenon was observed in many fields including telecommunication, hydrology, biomechanics, economy... A stationary second order process $Y = \{Y(k), k \in \mathbb{N}\}$ is said to be a LRD process if:

$$\sum_{k \in \mathbb{N}} |r_Y(k)| = \infty \text{ with } r_Y(k) = \mathbb{E}[Y(0)Y(k)].$$

Thus $Y(k)$ is depending on $Y(0)$ even if k is a very large lag. Another way for writing the LRD property is the following:

$$r_Y(k) \sim k^{2H-2} L(k), \text{ as } k \rightarrow \infty,$$

with $L(k)$ a slowly varying function (*i.e.* $\forall t > 0$, $L(xt)/L(x) \rightarrow 1$ when $x \rightarrow \infty$) and the Hurst parameter $H \in (\frac{1}{2}, 1)$.

The LRD is closely related to the self-similarity concept. A process $X = \{X(t), t \geq 0\}$ is so called a self-similar process with self-similarity exponent H , if $\forall c > 0$:

$$(X(ct))_t \stackrel{\mathcal{L}}{=} c^H (X(t))_t.$$

Now, if we consider the aggregated process $\{X(t), t \geq 0\}$ defined by $X(k) = \sum_{i=1}^k Y(i)$ with a LRD process Y , then under weak conditions (for instance Y is a Gaussian or a causal linear process), it can be proved that, roughly speaking, for $k \rightarrow \infty$, the distribution of $\{X(t), t \geq k\}$ is a self-similar distribution (see Doukhan *et al.*, 2003, for more details). The FGN is an example of a LRD Gaussian process. More precisely, $Y^H = \{Y^H(k), k \in \mathbb{N}\}$ is a FGN, when

$$r_{Y^H}(k) = \frac{\sigma^2}{2} (|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}) \quad \forall k \in \mathbb{N},$$

with $H \in (0, 1)$ and $\sigma^2 > 0$. As a consequence, for $H \in (\frac{1}{2}, 1)$, a usual Taylor formula implies

$$r_{Y^H}(k) \sim \sigma^2 H(2H-1)k^{2H-2}, \text{ when } k \rightarrow \infty.$$

For a zero-mean FGN, the corresponding aggregated process, denoted here X^H , is so-called the fractional Brownian motion (FBM) and X^H is a self-similar Gaussian process with self-similar parameter H and therefore satisfies,

$$\text{Var}(X^H(k)) = \sigma^2 |k|^{2H} \quad \forall k \in \mathbb{N}$$

(it can be even proved that X^H is the only Gaussian self-similar process with stationary increments). It is obvious that $Y^H(k) = X^H(k) - X^H(k-1)$, the sequence of the increments of a FBM, is a FGN.

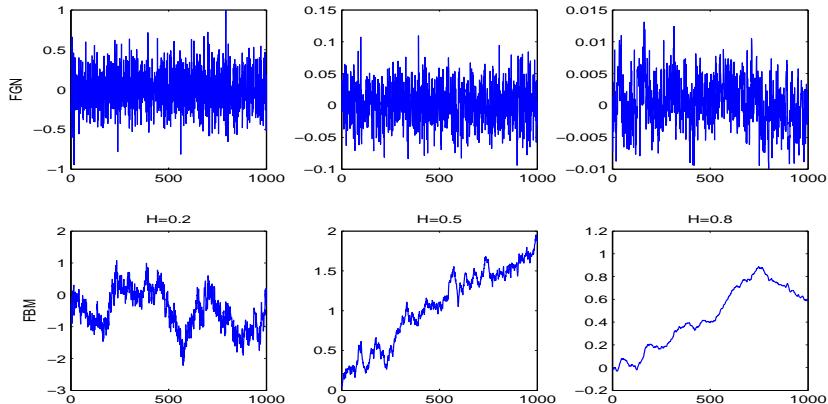


Fig. 5. Generated FGN trajectories and corresponding aggregated series (FBM) for $H = 0.2 < 0.5$ anti-persistent noise (left), $H = 0.5$ white noise (center) and $H = 0.8 > 0.5$ LRD process (right)

Several generated trajectories of FGN and corresponding FBM are presented in Fig. 5 for different values of H .

For testing if a HR path can be suitably model by a FGN, a first step consists in estimating H . Here we chose to use two estimators (but there exist many else, see for instance Doukhan *et al.*, 2003) that are known to be unchanged to the presence of a possible trend.

3.2. Two estimators of the Hurst parameter: DFA and wavelet based estimators

For estimating H , a frequently used method in the case of physiological data processing is the Detrended Fluctuation Analysis (DFA). The DFA method was introduced by Peng *et al.* Peng *et al.* (1994). The DFA method is a version for trended time series of the method of aggregated variance used for long-memory stationary process. It consists briefly on:

- Aggregated the process and divided it into windows with fixed length,
- Detrended the process from a linear regression in each windows,
- Computed the standard deviation of the residual errors (the DFA function) for all data,

(d) Estimated the coefficient of the power law from a log-log regression of the DFA function on the length of the chosen window (see Fig. 6).

After the first stage, the process is supposed to behave like a self-similar process with stationary increments added with a trend (see previously). The second stage is supposed to remove the trend. Finally, the third and fourth stages are the same than those of the aggregated method (for zero-mean stationary process). An example of the DFA method applied to a path of a FGN with different values of H is shown in Fig. 6.

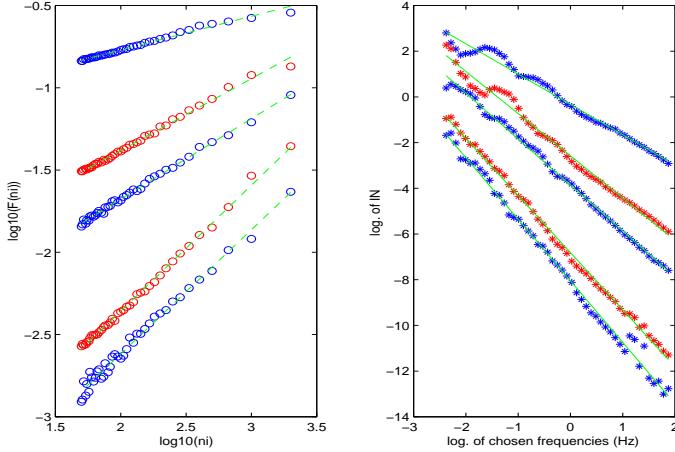


Fig. 6. Results of the DFA method and wavelet analysis applied to a path of a discretized FGN for different values of $H = 0.2, 0.4, 0.5, 0.7, 0.8$, with $N = 10000$

In Bardet and Kammoun (2008), the asymptotic properties of the DFA function in case of a FGN path $(Y(1), \dots, Y(N))$ are studied. In such a case the estimator \hat{H}_{DFA} converges to H with a non-optimal convergence rate ($N^{1/3}$ instead of $N^{1/2}$ reached for instance by maximum likelihood estimator). An extension of these results for a general class of stationary Gaussian LRD processes is also established. In this semiparametric frame, we have shown that the estimator \hat{H}_{DFA} converges to H with an optimal convergence rate (following the minimax criteria) when an optimal length of windows is known.

The processing of experimental data, and in particular physiological data, exhibits a major problem that is the nonstationarity of the signal. Hu *et al.* (2001) have studied different types of nonstationarity associated with examples of trends and deduced their effect on an added noise and the kind of competition who exists between this two signals. They have also explained (2002) the effects of three other types of nonstationarity, which are often encountered in real data. In Bardet and Kammoun (2008), we proved that \hat{H}_{DFA} does not converge to H when a polynomial trend (with degree greater or equal to 1) or a piecewise constant trend is added to a LRD process: the DFA method is clearly a non robust estimation of the Hurst parameter in case of trend.

For improving this estimation at least for polynomial trended LRD process, a wavelet based estimator is now considered. This method has been introduced by Flandrin (1992) and was

developed by Abry *et al.* (2003) and Bardet *et al.* (2000). In Wesfreid *et al.* (2005), a multifractal analysis of HR time series is presented for trying to unveil their scaling law behavior using the Wavelet Transform Modulus Maxima (WTMM) method.

Let $\psi : \mathbb{R} \rightarrow \mathbb{R}$ a function so-called the mother wavelet. Let $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ and denote $\lambda = (a, b)$. Then define the family of functions ψ_λ by

$$\psi_\lambda(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t}{a} - b\right)$$

Parameters a and b are so-called the scale and the shift of the wavelet transform. Let us underline that we consider a continuous wavelet transform. Let $d_Z(a, b)$ be the wavelet coefficient of the process $Z = \{Z(t), t \in \mathbb{R}\}$ for the scale a and the shift b , with

$$d_Z(a, b) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} \psi\left(\frac{t}{a} - b\right) Z(t) dt = \langle \psi_\lambda, Z \rangle_{L^2(\mathbb{R})}.$$

For a time series instead of a continuous time process, a Riemann sum can replace the previous integral for providing a discretized wavelet coefficient $e_Z(a, b)$. The function ψ is supposed to be a function such that it exists $M \in \mathbb{N}^*$ satisfying,

$$\int_{\mathbb{R}} t^m \psi(t) dt = 0 \quad \text{for all } m \in \{0, 1, \dots, M\}. \quad (2)$$

Therefore, ψ has its M first vanishing moments. The wavelet based method can be applied to LRD or self-similar processes for respectively estimating the Hurst or the self-similarity parameter. This method is based on the following properties: for Z a stationary LRD process or a self-similar process having stationary increments, for all $a > 0$, $(d_Z(a, b))_{b \in \mathbb{R}}$ is a zero-mean stationary process and

- If Z is a stationary LRD process,

$$\mathbb{E}(d_Z^2(a, b)) = \text{Var}(d_Z(a, b)) \sim C(\psi, H) a^{2H-1} \quad \text{when } a \rightarrow \infty$$

- If Z is a self-similar process having stationary increments,

$$\mathbb{E}(d_Z^2(a, b)) = \text{Var}(d_Z(a, b)) \sim K(\psi, H) a^{2H+1} \quad \text{for all } a > 0$$

with $C(\psi, H)$ and $K(\psi, H)$ two positive constants depending only on ψ and H (those results are proved in Flandrin, 1992, Abry *et al.*, 1998). Therefore, in both these cases, the variance of wavelet coefficients is a power law of a , and a log-log regression provides an estimator of H . From a path $(Z(1), \dots, Z(N))$, the estimator will be deduced from the log-log regression of the "natural" sample variance of discretized wavelet coefficients, *i.e.*,

$$S_N(a) = \frac{1}{[N/a]} \sum_{i=1}^{[N/a]} e_Z^2(a, i). \quad (3)$$

A graph $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$ is drawn from a priori family of scales and the slope of the least square regression line provides the estimator \hat{H}_{WAV} of H . In the semiparametric frame of a general class of stationary Gaussian LRD processes (more general than the previous

Table 1. Comparison of the two samples of estimations of H with 100 realizations of fGn path ($N=10000$) with DFA and wavelets methods (The p -val was deduced from comparison of the mean of each sample with theoretical one at the 5% level)

H_{fGn}	$ Bias_{\hat{H}_{DFA}} $	$ Bias_{\hat{H}_{WAV}} $	$p\text{-val}_{DFA}$	$p\text{-val}_{WAV}$	\sqrt{MSE}_{DFA}	\sqrt{MSE}_{WAV}
0.50	0.0064	0.0071	0.0152	0.0983	0.0271	0.0433
0.60	0.0092	0.0009	0.0017	0.8289	0.0304	0.0405
0.70	0.0141	0.0015	10^{-5}	0.7342	0.0347	0.0436
0.80	0.0125	0.0050	0.0002	0.1978	0.0349	0.0391
0.90	0.0179	0.0062	$2 \cdot 10^{-6}$	0.1030	0.0407	0.0448

semiparametric context required for \hat{H}_{DFA}), it was established by Moulines *et al.* (2007) that the estimator \hat{H}_{WAV} converges to H with an optimal convergence rate (following the minimax criteria) when an optimal length of windows is known. Thus, theoretical asymptotic behaviors of \hat{H}_{DFA} and \hat{H}_{WAV} are comparable for FGN and a semiparametrical class of LRD Gaussian processes (however more general for \hat{H}_{WAV}).

This is not true any more when a polynomial trended LRD (or self-similar) processes is considered. Indeed, Abry *et al.* (1998) remarked that every degree M polynomial trend is without effects on \hat{H}_{WAV} since ψ has its M first vanishing moments. Therefore, the larger M , the more robust \hat{H}_{WAV} is.

Finally, Bardet (2002) established a chi-squared goodness-of-fit test for a path of FBM (therefore for aggregated FGN) using wavelet analysis. This test is based on a (penalized) distance between the points $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$ and a pseudo-generalized least square regression line (here the scales a_i are selected to behave as $N^{1/3}$).

In the Table 1 appear the different estimations of H computed from the DFA and wavelet analysis methods for 100 realizations of FGN paths with $N = 10000$. We choose for these simulations the concrete procedure of wavelet analysis developed by Abry *et al.* (2003) (a Daubechies wavelet is chosen and a Mallat's fast pyramidal algorithm is used to compute wavelet coefficients). In one hand, the wavelets method appear slightly more effective than DFA method considering the p-value which is very low for the sample of the DFA estimations compared to wavelet analysis estimations. This is essentially due to the estimator bias which is more important in the case of DFA. In the other hand, if we consider the root of MSE which is the sum of the squared bias and the variance, the DFA estimator seems to be slightly more effective. Note that for FGN processes (without trend), the Whittle maximum likelihood estimator of H gives a "better" results (see Taqqu *et al.*, 1999).

3.3. Application of both the estimators to HR data

Both these estimators of H can also be applied to the HR time series of the 9 athletes. The following figures Fig. 7 and Fig. 8 exhibit examples of applications of both the estimation method to HR data. For each athlete, it was first done to the whole time series, and then to the different phases of the race (as it was obtained from the detection of abrupt changes, see Section 2). The estimation results of H , for the different signals observed during the three phases of the race, are recapitulated in the Table 2 using wavelets method and in Table 3 using DFA method.

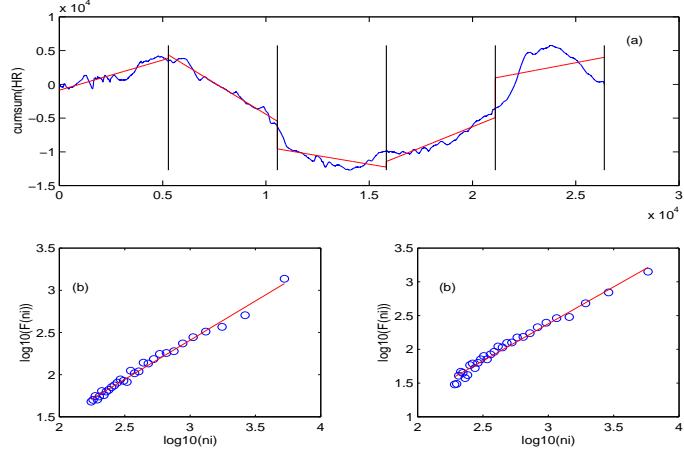


Fig. 7. Two first steps of the DFA method applied to a HR series (up) and results of the DFA method applied to HR series for two different athletes (down)

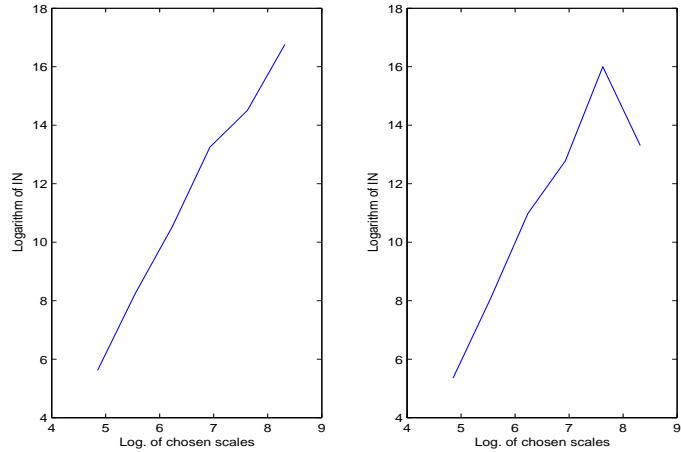


Fig. 8. The log-log graph of the variance of wavelet coefficients relating to the HR series observed during the race and in the end of race (Ath2)

Two main problems result from these different estimations. First, \hat{H}_{DFA} and \hat{H}_{WAV} are often larger than 1. However, the FGN is only defined for $H \in (0, 1)$. For defining a process allowing $H > 1$, three main assumptions of FGN have to be changed:

- (a) the assumption that the process is a stationary process;
- (b) the assumption that the process is a Gaussian process;
- (c) the assumption that only two parameters (H and σ^2) are sufficient to define the process.

Table 2. Estimated H with wavelets methods for HR series of different athletes

	HR series	Phases		
		Beginning	Middle	Race end
Ath1	0.8931	1.1268	1.1064	1.2773
Ath2	1.1174	0.7871	1.0916	0.8472
Ath3	1.0208	1.0315	1.1797	-
Ath4	0.9273	-	1.0407	0.7925
Ath5	1.0986	1.3110	1.0113	1.3952
Ath6	1.0769	1.5020	1.1597	1.3673
Ath7	1.0654	1.4237	1.1766	1.0151
Ath8	0.9568	1.6600	0.9699	1.1948
Ath9	0.9379	1.5791	0.9877	0.7263

In the sequel (see below), a new model is proposed. Both the first assumptions are still satisfied and the third one is replaced by a semiparametric assumption.

The second problem is implied by the results of the goodness-of-fit test (for wavelet analysis method). Indeed, this test is never accepted as well for the whole time series as for the partial times series. An explanation of such a phenomenon can be deduced from Figure 8: for the wavelet analysis, the points $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$ are clearly lined for $a_i \leq a_m$, but not exactly lined for $a_i \geq a_m$. Thus the HR time series seems to nearly behave like a FGN for "small" scales (or high frequencies), but not for "large" scales (or small frequencies). A process following this conclusion can not be the better fit of HR time series...

Remark: this last conclusion leads also to a clear advantage of wavelet based over DFA estimator. Indeed, the DFA algorithm measures only one exponent characterizing the entire signal. Then, this method corresponds rather to the study of "monofractal" signals such as FGN. At the contrary, the wavelet method provides the graph $(\log a_i, \log S_N(a_i))_{1 \leq i \leq \ell}$ which can be very interesting for analyze the multifractal behavior of data (see also Billat *et al.*, 2005).

4. A new model for modeling HR data: a locally fractional Gaussian noise

4.1. The locally fractional Gaussian noise

In Bardet and Bertrand (2007), a generalization of the FBM, so-called the (M_K) -multiscale FBM, was introduced. The (M_0) -FBM is a FBM with self-similarity parameter H_0 . Roughly speaking, the (M_K) -FBM has the same harmonizable representation (and therefore quite the same behavior as the FBM) than a FBM with self-similarity parameter H_i for frequencies $|\xi| \in [\omega_i, \omega_{i+1}]$ for all $i = 0, \dots, K$ ($K \in \mathbb{N}$). For instance, a (M_1) -FBM behaves as a FBM with self-similarity parameter H_0 for small frequencies and as a FBM with self-similarity parameter H_1 for high frequencies. Such a model was fruitfully used for modeling biomechanical signals (position of the center of pressure on a force platform during quiet postural stance measured at a frequency of 100 Hz for the one minute period).

Here, Fig. 9 suggests than a fitted model for aggregated HR data should behave like a FBM with self-similarity parameter H for low frequencies and differently for high frequencies

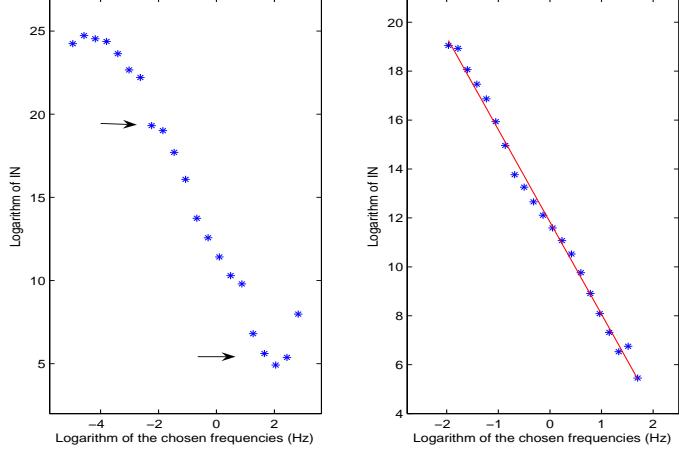


Fig. 9. The log-log graph of the variance of wavelet coefficients relating to the HR series observed during the arrival phase (Ath6) with a frequency band of [0.01 12](right) and of [0.2 4](left).

(and not necessary like a FBM). Thus define a locally fractional Brownian motion $X_\rho = \{X_\rho(t), t \in \mathbb{R}\}$ as the process such that:

$$X_\rho(t) = \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{\rho(\xi)} \widehat{W}(d\xi)$$

where the function $\rho : \mathbb{R} \rightarrow [0, \infty)$ is an even continuous function such that:

- $\rho(\xi) = \frac{1}{\sigma} |\xi|^{H+1/2}$ for $|\xi| \in [\omega_0, \omega_1]$ with $H \in \mathbb{R}$, $\sigma > 0$ and $0 < \omega_0 < \omega_1$
- $\int_{\mathbb{R}} (1 \wedge |\xi|^2) \frac{1}{\rho^2(\xi)} d\xi < \infty$.

and $W(d\xi)$ is a Brownian measure and $\widehat{W}(d\xi)$ its Fourier transform in the distribution meaning. Cramér and Leadbetter (1967) proved the existence of such Gaussian process with stationary increments. The main advantages of such process compared to usual FBM are the following:

- (a) X_ρ "behaves" like a FBM only for local band of frequencies;
- (b) In this band, the parameter H is not restricted to be in $(0, 1)$: it is in \mathbb{R} .

From this definition, one deduces a possible model for HR data:

$$Y_\rho(t) = X_\rho(t+1) - X_\rho(t) = 2 \cdot \mathcal{R}e \left(\int_{\mathbb{R}} \frac{e^{it\xi} \sin(\xi/2)}{\rho(\xi)} \widehat{W}(d\xi) \right) \text{ for } t \in \mathbb{R}.$$

Note that $Y_\rho = \{Y_\rho(t), t \in \mathbb{R}\}$ is a stationary Gaussian process and the function $2 \sin(\xi/2) \rho^{-1}(\xi)$ is so-called the spectral density of Y_ρ .

Let $\Delta_N \rightarrow 0$ and $N\Delta_N \rightarrow \infty$ when $N \rightarrow \infty$. The wavelet based estimator can provide a convergent estimation of H when a path

$$(Y_\rho(\Delta_N), Y_\rho(2\Delta_N), \dots, Y_\rho(N\Delta_N))$$

and therefore a path $(X_\rho(\Delta_N), \dots, X_\rho(N\Delta_N))$ is observed. Indeed, consider a "mother" wavelet ψ such that $\psi : \mathbb{R} \mapsto \mathbb{R}$ is a \mathcal{C}^∞ function satisfying :

- for all $s \geq 0$, $\int_{\mathbb{R}} |t^s \psi(t)| dt < \infty$;
- its Fourier transform $\widehat{\psi}(\xi)$ is an even function compactly supported on $[-\beta, -\alpha] \cup [\alpha, \beta]$ with $0 < \alpha < \beta$.

Then, using results of Bardet and Bertrand (2007), for all $a > 0$ such that $[\frac{\alpha}{a}, \frac{\beta}{a}] \subset [\omega_0, \omega_1]$, i.e. $a \in [\frac{\beta}{\omega_1}, \frac{\alpha}{\omega_0}]$, $(d_{X_\rho}(a, b))_{b \in \mathbb{R}}$ is a stationary Gaussian process and

$$\mathbb{E}(d_{X_\rho}^2(a, .)) = K(\psi, H, \sigma) \cdot a^{2H+1},$$

with $K(\psi, H, \sigma) > 0$ only depending on ψ, H and σ . However this property is checked if and only if the function ψ is chosen such that:

$$\frac{\beta}{\alpha} < \frac{\omega_1}{\omega_0}.$$

Moreover, for $a \in [\frac{\beta}{\omega_1}, \frac{\alpha}{\omega_0}]$, the sample variance $S_N(a)$ defined in (3) and computed from a path $(X_\rho(\Delta_N), \dots, X_\rho(N\Delta_N))$ converges to $\mathbb{E}(d_{X_\rho}^2(a, .))$ and satisfies a central limit theorem with convergence rate $\sqrt{N\Delta_N}$. Thus, with fixed scales $(a_1, \dots, a_\ell) \in [\frac{\beta}{\omega_1}, \frac{\alpha}{\omega_0}]^\ell$, a log-log-regression of $(a_i, S_N(a_i))_{1 \leq i \leq \ell}$ provides an estimation of H (and a central limit theorem with convergence rate $N\Delta_N$ satisfied by \widehat{H}_{WAV} can also be established). As previously, we consider also chi-squared goodness-of-fit test based on the wavelet analysis and defined as a weighted distance between points $(\log(a_i), \log(S_N(a_i)))_{1 \leq i \leq \ell}$ and a pseudo-generalized regression line.

Remark: The main problem with these estimator and test is the localization of the suitable frequency band $[\omega_0, \omega_1]$ (ω_0 and ω_1 are assumed to be unknown parameters). A solution consists in selecting a very large band of scales and determining then graphically the "most" linear part of the set of points $(\log(a_i), \log(S_N(a_i)))_{1 \leq i \leq \ell}$. Another possible way may be to compute an adaptive estimator of this band using a quadratic criterion (following a similar procedure than in Bardet and Bertrand, 2007). Here, like 9 different paths of HR data are observed, a common frequency band $[\overline{\omega_0}, \overline{\omega_1}]$ can be graphically obtained and used for whole HR data (see above).

4.2. Application to HR data

First, one considers that a HR time series $(Y(1), \dots, Y(n))$ can be written $(Y_\rho(\Delta_n), Y_\rho(2\Delta_n), \dots, Y_\rho(n\Delta_n))$, $Y_\rho = \{Y_\rho(t), t \in \mathbb{R}\}$ a process defined as previously. Secondly, the wavelet

Table 3. Estimated \hat{H} , with DFA and wavelets methods, for HR series of different athletes

(*) The series for which the test is rejected. Comparison of the two samples $(\hat{H}_{DFA})_{1,\dots,9}$ and $(\hat{H}_{WAV})_{1,\dots,9}$ for whole and partial series (p-value)

	HR series		Race beginning		During the race		End of race	
	\hat{H}_{DFA}	\hat{H}_{WAV}	\hat{H}_{DFA}	\hat{H}_{WAV}	\hat{H}_{DFA}	\hat{H}_{WAV}	\hat{H}_{DFA}	\hat{H}_{WAV}
Ath1	0.928	1.288*	1.032	1.192	1.060	1.214	0.429	1.400
Ath2	1.095	1.268*	0.905	0.973	1.126	1.108	1.240	1.452*
Ath3	1.163	1.048	0.553	0.898	1.130	1.172	-	-
Ath4	1.193	0.916*	-	-	1.098	1.249*	1.172	1.260
Ath5	1.239	1.110	1.267	1.117*	1.133	1.205	1.273	1.348*
Ath6	1.247	1.084*	1.237	1.106	1.091	1.172	1.436	1.338
Ath7	1.155	1.095	0.850	1.295	1.182	1.186*	1.129	1.209
Ath8	1.258	1.011	1.304	1.128*	0.995	1.134	1.122	1.247
Ath9	1.243	1.429*	0.820	1.019	1.127	1.535*	1.250	1.238*
<i>p</i> -value	0.6414		0.3723		0.0225		0.1260	
<i>F</i> -stat	0.23		0.85		6.38		2.65	

analysis is applied to the 9 (whole or partial) HR time series (the chosen "mother" wavelet is a kind of Lemarié-Meyer wavelet such that $\beta = 2\alpha$). Using first a very large band of scales for all HR time series (for example $[0.01, 12]$ in Fig. 10), one estimation of frequency band is deduced: $[\bar{\omega}_0, \bar{\omega}_1] = [0.2, 4]$ is the chosen frequency band for the whole and partial signals.

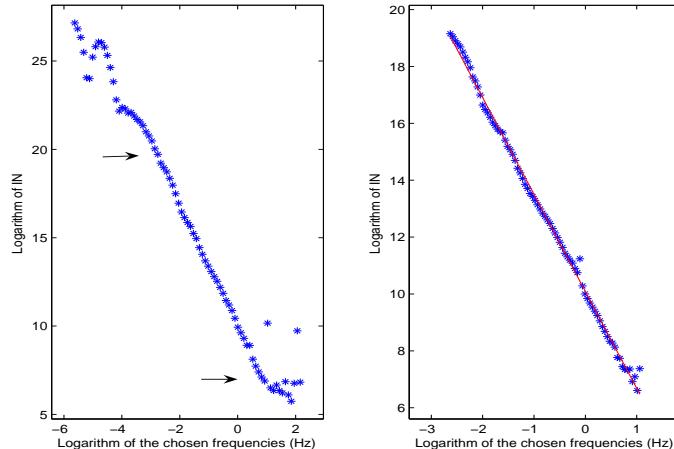


Fig. 10. The log-log graph of the variance of wavelet coefficients relating to the HR series observed in the middle of the exercise (Ath5)

The estimation results of H , for the different signals observed during the three phases of the race, are recapitulated in the Table 3.

Both DFA and wavelet analysis methods provide estimations of Hurst exponent which reflect the possible modeling of HR data with long range dependence time series.

We also note that with a p-value of 0.64, both the samples $(\hat{H}_{DFA})_{1,\dots,9}$ and $(\hat{H}_{WAV})_{1,\dots,9}$ obtained from all HR time series are significantly close.

The same comparison can also be done when the three characteristic stages of the race (beginning, middle and end of the race) are distinguished. The result is different. Indeed, the corresponding p-values between $(\hat{H}_{DFA})_{1,\dots,9}$ and $(\hat{H}_{WAV})_{1,\dots,9}$ are significatively different in the middle part of the race (and relatively different in the stage of race end).

In spite of values relating to the estimator of H for all the athletes in the different phases which are relatively large, the DFA has sometimes tendency to under estimating this parameter like in the race beginning (Ath3) and the end of race (Ath1). Indeed, these value are clearly due to a certain trend supports by the fact that data points in log-log plot (Fig. 11) have not a straight line form, and we have proved in Bardet and Kammoun (2008) that the DFA method is not robust in the case of trended long range dependent process. However in both the cases, the wavelets method is more effective since it removes sufficiently this kind of trend.

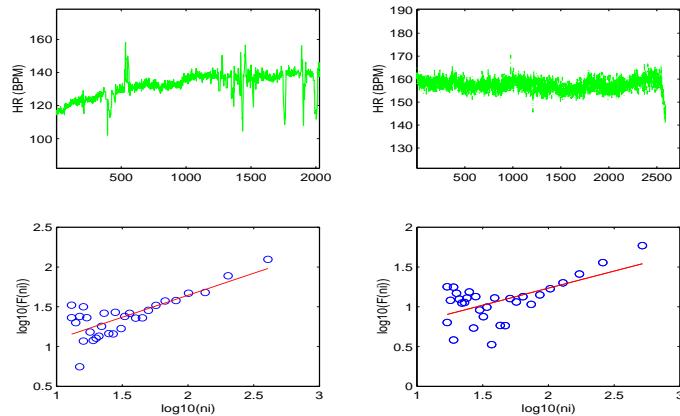


Fig. 11. The results of the DFA method applied to records for race beginning (Ath3) (left) and for end of race (Ath1) (right)

For HR data and when the goodness-of-fit test is accepted, the wavelet method shows a fractal parameter H close to 1. According to the different studies (using DFA method) about physiologic time series for distinguishing healthy from pathologic data sets (see Goldberger *et al.* (2002), Peng *et al.* (1995), Peng *et al.* (1993)), an exponent $H \simeq 1$ indicate a healthy cardiac HR time series. Indeed, for the study concerning a 24 hours recorded interbeat time series during the exercise for healthy adults and heart failure adults, the following results are obtained: for healthy subjects, $H = 1.01 \pm 0.16$, for the group of heart failure subjects $H = 1.24 \pm 0.22$.

During the different stages of the marathon race, a small increase of the fractal parameter H is observed especially at the end of races. This behavior and this evolution may be associated with fatigue appearing during the last phase of the marathon. This evolution can not be observed with DFA method. Indeed, in one hand, when we observe the three 9-samples of wavelet estimators (related to the 3 phases of the race), the p-value (see Fig.

12) indicates a significantly difference due precisely to this evolution of the fractal parameter. On the other hand, a large p-value (0.85) is obtained for the same test using DFA estimation.

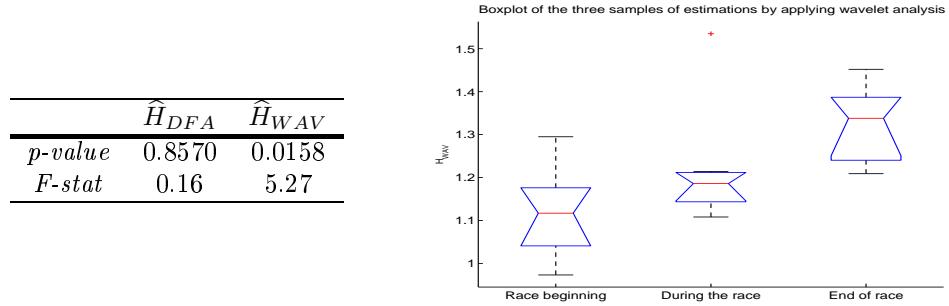


Fig. 12. Comparison of the three samples constituting by estimations in the beginning of race, during the race and then in the race end by the DFA and wavelet methods

The representation given by Fig. 12, highlight a difference in the behaviors of HR series in the beginning of the race and in the end of race. Indeed, the dispersions in the first and last sample are more important than in the middle of race and it seems that each athlete starts and finished the race at his own rhythm but in the middle athletes seems to have the same rate.

5. Conclusion

As indicated in the beginning of the last section, our main goal is to see whether the heart rate time series during the race have specific properties that of scaling law behavior. The wavelet analysis and the DFA methods are applied to 9 HR time series during the whole and also the different three phases of the race (beginning, middle and end of race) obtained by an automatic procedure. Even if their results are not exactly the same, both methods provide Hurst exponents which reflect the possible modeling of HR data by a LRD time series. However, in Bardet and Kammoun (2008), even if the DFA estimator of Hurst parameter is proved to be convergent with a reasonable convergence rate for LRD stationary Gaussian processes, it is not at all a robust method in case of trend. The wavelet based method provides a more precise and robust estimator of the Hurst parameter. Thus, the results obtained from this wavelet estimator seem to be more valid.

Moreover, a chi-squared goodness-of-fit test can also be deduced from this method. It seems to show that a classical LRD stationary Gaussian process is not exactly a suitable model for HR data. Graphs obtained with wavelet analysis also show that a locally fractional Gaussian noise, a semiparametric process defined in Section 3 could be more relevant to model these data. A chi-squared test confirms the goodness-of-fit of such a model. Thus, using the wavelet estimation of a fractal parameter in a specific frequency band, one obtains a conclusion relatively close to those obtained by other studies (conclusion which can not be detected with DFA method): these fractal parameters increase through the race phases, what may be explained with fatigue appearing during the last phase of the marathon. Thus

this fractal parameter may be a relevant factor to detect a change during a long-distance race.

Finally, for the 9 athletes and as the test is validated with significance level around 0.65, we can estimate $\hat{H}_{beginning}$ at 1.1, the \hat{H}_{middle} at 1.2 and \hat{H}_{end} at 1.3 with a larger confidence interval at the beginning and the end of the race. This behavior could bring a new way of understanding what is happening during a race.

References

Abry, P., Flandrin, P., Taqqu, M. and Veitch, D. (2003) Self-similarity and long-range dependence through the wavelets lens. *Theory and applications of Long-Range Dependence*, 527-556, Birkhäuser.

Abry, P., Veitch, D. and Flandrin, P. (1998) Long-range dependent: revisiting aggregation with wavelets. *J. Time Ser. Anal.*, 253-266.

Absil, P.A., Sepulchre, R., Bilge, A. and Gérard, P. (1999) Nonlinear analysis of cardiac rhythm fluctuations using DFA method. *J. Physica A: Statistical mechanics and its applications*, 235-244.

Bagger, M., Petersen, P.H. and Pedersen, P.K. (2003) Biological Variation in variables associated with exercise training. *Int. J. Sports Med.*, 433-440.

Bai, J. (1998) Least squares estimation of a shift in linear processes. *J. of Time Series Anal.*, 5, 453-472.

Bai, J. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47-78.

Bardet, J.M. (2002) Statistical study of the wavelet analysis of fractional Brownian motion. *IEEE Trans. Inform. Theory.*, 48, 991-999.

Bardet, J.M., Bertrand, P. (2007) Identification of the multiscale fractional Brownian motion with biomechanical applications. *Journal of Time Series Analysis*, 28, 1-52.

Bardet, J.M. and Kammoun, I. (2008) Asymptotic Properties of the Detrended Fluctuation Analysis of Long Range Dependence Processes. *To appear in IEEE Trans. Inform. Theory.*

Bardet, J.M., Lang, G., Moulines, E. and Soulier, P. (2000) Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes*, 3, 85-99.

Bardet, J.M., Lang, G., Oppenheim, G., Philippe, A. and Taqqu, M.S. (2003) Generators of long-range dependent processes : A survey. *Long-range Dependence : Theory and Applications*, Birkhauser.

Billat, V., Wesfreid, E., Cottin, F., Kapfer, C. and Koralztein, J.P. (2003) Fractal analysis of speed and physiological oscillations in long- and middle-distance running: Effect of training. *International Journal of Computer Science in Sport*, 16-30. <http://www.billat.net/>.

Flandrin, P. (1992) Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Inform. Theory*, 38, 910-917.

Goldberger, A.L. (2001) Heartbeats, hormones, and health : is variability the spice of life? *Am J Respir Crit Care Med*, 163, 1289-1290.

Goldberger, A.L., Amaral, L.A.N., Hausdorff, J.M., Ivanov, P.C., Peng, C.K. and Stanley, H.E. (2002) Fractal Dynamics in Physiology: Alterations with Disease and Aging. *PNAS*, 99, 2466-2472.

Lavielle, M. (1999) Detection of multiple changes in a sequence of random variables. *Stoch. Process Appl*, 79-102. <http://www.math.u-psud.fr/~lavielle/>.

Lavielle, M. and Moulines, E. (2000) Least-squares estimation of an unknown number of shifts in a time series. *J. of Time Series Anal.*, 21, 33-59.

Moulines, E., Roueff, F. and Taqqu, M.S. (2007), On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *J. Time Ser. Anal.*, 28, 155-187.

Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In P. Doukhan, G. Oppenheim and M.S. Taqqu editors. *Theory and applications of long-range dependence*, 251-301, Birkhäuser, Boston.

Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E. and Goldberger, A.L. (1994) Mosaic organization of DNA nucleotides. *Physical Review E*, 49, 1685-1689.

Peng, C.K., Havlin, S., Stanley, H.E. and Goldberger, A.L. (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5, 82.

Peng, C.K., Mietus, J., Hausdorff, J., Havlin, S., Stanley H.E. and Goldberger A.L. Long-Range Anticorrelations and Non-Gaussian Behavior of the Heartbeat. *Phys. Rev. Lett.*, 70, 1343-1346.

Pikkujämsä, S.M., Mäkkilä, T.H., Airaksinen, K.E. and Huikuri, H.V. (2001) Determinants and interindividual variation of R-R interval dynamics in healthy middle-aged subjects. *Am. J. Physiol. Heart Circ. Physiol.*, 280, 1400-1406.

Samorodnitsky, G. and Taqqu, M.S. (1994) Stable non-Gaussian Random Processes. *Chapman and Hall*.

Stoev, S., Taqqu, M.S., Park, C. and Marron, J.S. (2005) On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of Internet traffic. *Elsevier*, 49, 423-445.

Taqqu, M.S. and Teverovsky, V. (1997) Robustness of whittle type estimators for time series with long range dependence. *J. Stochastic Model*, 13, 723-757.

Taqqu, M.S., Teverovsky, V. and Willinger, W. (1995) Estimators for long-range dependence: an empirical study. *Fractals*, 3, 785-798.

Wesfreid, E., Billat, V. and Meyer, Y. (2005) Multifractal analysis of heartbeat time series in human races. *Appl. Comput. Harmon. Anal.*, 18, 329-335.